

WHAT IS CLAIMED IS:

1. A method for identifying at least two component words in a compound word, comprising:
 - receiving the compound word;
 - searching a lexicon for the compound word; and
 - identifying the component words for the compound word based on entries contained in the lexicon.
2. The method of claim 1 wherein identifying the component words further comprises:
 - matching a first portion of the compound word with a first entry in the lexicon; and
 - matching a second portion of the compound word with a second entry in the lexicon.
3. The method of claim 2 further comprising:
 - matching additional portions of the compound word until reaching a final character of the compound word.
4. The method of claim 3 further comprising:
 - generating a first hypothesis of the component words of the compound word, wherein the first hypothesis is a combination of the matched portions of the compound word.
5. The method of claim 4 further comprising:
 - generating additional hypotheses of the component words of the compound word.

6. The method of claim 5 further comprising:
ranking each identified hypothesis based on a
relative likelihood of being a correct
representation of the component words of
the compound word.
7. The method of claim 6 wherein ranking is based
on data obtained through statistical analysis
8. The method of claim 5 wherein matching the first
portion comprises matching characters in the compound
word starting from a first character of the compound
word; and
wherein matching the second portion comprises
matching characters in the compound word
starting from a first character that
follows a last character of the first
portion.
9. The method of claim 8 wherein matching the first
portion further comprising:
searching the lexicon for an entry that matches
the first portion;
if a match is found;
analyzing the second portion for matches
with entries in the lexicon;
if a match is found for the second portion;
generating a hypotheses in a list of
hypothesis, wherein the

hypothesis is a combination of the first portion and the second portion.

10. The method of claim 9 wherein if additional matches are found for the first and second portions in the lexicon, adding these matches as alternative hypotheses to the list of hypotheses

11. The method of claim 9 further comprising:
adding the hypothesis to the list of hypotheses regardless of whether the entry includes the segment indication; and
excluding a resulting trace as invalid at a final evaluation of hypotheses.

12. The method of claim 11 wherein if the second portion includes a portion which is a non-final portion further comprising the steps of:
checking if the non-final portion includes the segment indication;
adding the hypothesis to the list of hypotheses;
and
excluding the resulting trace as invalid at the final evaluation of hypotheses, if the anti-seg bit is present in the final segment of the second portion..

13. The method of claim 9 further comprising:
returning to the first portion;

adding a character following the last character
in the first portion to the first portion;
repeating the steps of searching, generating,
and analyzing;
if additional matches are found for the first
and second portions in the lexicon, adding
these matches as alternative hypotheses to
the list of hypotheses

14. The method of claim 9 further comprising:
prior to adding the hypothesis to the list of
hypotheses, checking if the matching entry
in the lexicon for the first portion
includes a segment indication; and
adding the hypothesis to the list of hypotheses
only if the entry includes the segment
indication.

15. The method of claim 14 wherein if the second
portion includes a portion which is a non-final
portion, checking if the non-final portion includes
the segment indication; and
adding the hypothesis to the list of hypotheses
only if the entries for all non-final
segments include the segment indication.

16. The method of claim 9 further comprising:
prior to adding the hypothesis to the list of
hypotheses, checking if the matching entry
in the lexicon for the second portion or a

last portion of the second portion includes an anti-segment indication; and adding the string to the hypothesis only if the entry does not include the anti-segment indication.

17. A computer readable medium containing computer executable instructions that, when executed, cause a computer to perform the steps of:

receiving the compound word;
searching a lexicon for the compound word;
identifying the component words for the compound word based on entries contained in the lexicon by matching a first portion of the compound word with a first entry in the lexicon; and
matching a second portion of the compound word with a second entry in the lexicon.

18. The computer readable medium of claim 17 further comprising instructions to perform the steps of:

matching additional portions of the compound word until reaching a final character of the compound word.

19. The computer readable medium of claim 18 further comprising instructions to perform the steps of:

generating a first hypothesis of the component words of the compound word, wherein the

first hypothesis is a combination of the matched portions of the compound word; and generating additional hypotheses of the component words of the compound word if additional matched portions are found in the compound word.

20. The computer readable medium of claim 19 further comprising instructions to perform the steps of:

ranking each identified hypothesis based on a relative likelihood of being a correct representation of the component words of the compound word.

21. The computer readable medium of claim 17 further comprising instructions to perform the steps of:

wherein matching the first portion comprises matching characters in the compound word starting from a first character of the compound word; and

wherein matching the second portion comprises matching characters in the compound word starting from a first character that follows a last character of the first portion.

22. The computer readable medium of claim 21 wherein matching the first portion further comprises instructions to perform the steps of:

searching the lexicon for an entry that matches
the first portion;
if a match is found;
analyzing the second portion for matches
with entries in the lexicon;
if a match is found for the second portion;
generating a hypotheses in a list of
hypothesis, wherein the
hypothesis is a combination of
the first portion and the second
portion.

23. The computer readable medium of claim 22 wherein
if additional matches are found for the first and
second portions in the lexicon, adding these matches
as alternative hypotheses to the list of hypotheses

24. The computer readable medium of claim 23 further
comprising instructions to perform the steps of:
excluding a resulting hypothesis as invalid at a
final evaluation of hypotheses.

25. The computer readable medium of claim 24 wherein
if the second portion includes a portion which is a
non-final portion further comprises in instructions
to perform the steps of:
checking if the non-final portion includes the
segment indication;
adding the hypothesis to the list of hypotheses;
and

excluding the resulting hypothesis as invalid at the final evaluation of hypotheses, if the anti-seg bit is present in the final segment of the second portion.

26. A method of spell-checking a compound word, comprising:

searching a lexicon for the compound word;
if the compound word is not found in the lexicon;
identifying component words comprising the compound word;
comparing each of the identified component words with entries in the lexicon;
determining if each of the component words is used correctly; and
if at least one of the identified component words is used incorrectly indicating that the compound word includes a spelling error.

27. The method of claim 26 further comprising:
generating a spelling suggestion for the compound word based on the component words.

28. The method of claim 27 wherein determining if each component word is used correctly comprises:

if the component word is not the last component word in the compound, checking if the component word has a segment identifier;

if the component word does not have the segment identifier applying spelling transition rules to the component word; and
rechecking if the new component word is in the lexicon with the segment identifier.

29. The method of claim 28 wherein if applying speller transition rules results in too many unwanted lexical matches, further comprising the possibility of:

searching the entries in the lexicon for the component word to identify variations that include the segment identifier;
generating new compound words that include the identified variations of the component word; and
presenting those new compound words as the spelling suggestion to the user.

30. The method of claim 26 wherein determining if each component word is used correctly comprises:

if the component word is the last component word in the compound, checking if the component word has an anti-segment identifier;
if the component word has the anti-segment identifier searching the entry in the lexicon for the component word to identify variations that do not include the anti-segment identifier; and

if a variation of the component word does not include the anti-segment identifier, suggesting that variation of the component word as the spelling suggestion for the compound word.

31. The method of claim 28 wherein applying speller transition rules in the compound word includes adding additional characters to the component word that are not present in the compound word.

32. The method of claim 28 wherein applying speller transition rules in the compound word includes changing characters comprising the component word.

33. The method of claim 28 wherein adding an additional character via speller transition rules to the component word adds no more than two additional characters to the component word.

34. The method of claim 29 wherein presenting spelling suggestions presents the spelling suggestions only when applying speller transition rules adds, removes, or changes no more than two characters to the word.

35. The method of claim 29 wherein presenting spelling suggestions only presents spelling suggestions if no more than two component words in the compound word required correction.